

Real-time Surveillance Application by Multiple Detectors and Compressive Trackers

Jihan Li^{1,2}, Renjie Li^{1,2}, Raúl Mohedano², Carlos R. del-Blanco², Carlos Cuevas², Guillermo Gallego², Narciso García

¹ Beijing University of Posts and Telecommunications (China), ² Universidad Politécnica de Madrid (Spain).

{jli, rli, rmp, cda, ccr, ggb, narciso}@gti.ssr.upm.es

Abstract — *A real-time surveillance system for IP network cameras is presented. Motion, part-body, and whole-body detectors are efficiently combined to generate robust and fast detections, which feed multiple compressive trackers. The generated trajectories are then improved using a re-identification strategy for long term operation.*

Index Terms — Real-time Pedestrian Detection and Tracking, Part-body Detection, Compressive Tracking, Re-identification

I. INTRODUCTION

Smart surveillance systems are becoming essential, not only in security-sensitive environments such as banks, department stores and factories, but also in offices and domestic environments. However, traditionally, video data has been mainly used as a forensic tool, making no use of the online potential of visual surveillance [1]. Recently, there has been an increasing interest in developing new real-time techniques for unveiling such potential in scene understanding.

Effective online surveillance imposes strong requirements on the pedestrian tracking modules. Long-term robustness cannot be ensured by traditional tracking algorithms alone. Classifiers for both pedestrian detection and re-identification are required. However, online real-time performance prevents from using offline re-training schemes to increase the classification performance. One possible solution is to consider the inclusion of classifiers as a supporting step in the long-term tracking for re-identification and feedback purposes, while using traditional tracking approaches in the short-term to fulfil the required online restrictions.

Regarding the detection techniques based on classification, many of them use Histograms of Oriented Gradients (HOG) as feature descriptor. However, they usually obtain satisfactory results when the whole human body appears in the scene, failing in more complex scenes, where only some parts of the body are visible [2]. As for tracking, it is common to use algorithms based on Mean-Shift or Particle Filtering along with a color-based appearance model, but their accuracy is largely compromised in situations where the scene background and human body have similar color distributions [3]. Regarding pedestrian re-identification, histogram matching is a

standard despite the fact that it only takes into account color information, discarding others types of features.

In this paper, we propose a real-time pedestrian detection and tracking system for IP network surveillance cameras. After applying motion detection by background subtraction [4], potential areas with moving pedestrians are obtained. A whole-body HOG-based detector is used along with a part-body detector to reduce the miss rate and deal with occlusions. A compressive approach is adopted in the tracking stage [5], achieving real-time processing. The fingerprint of each pedestrian provides sufficient information to distinguish it from the rest. This system can be easily installed in multiple types of environment since it is suitable for IP cameras and shows promising results in terms of efficiency and robustness.

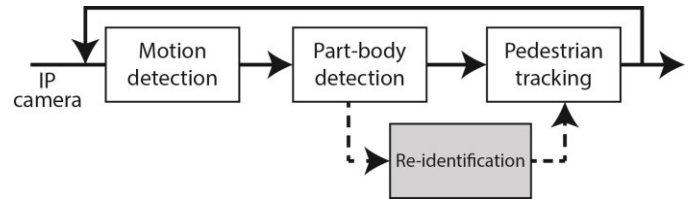


Fig. 1. High-level block diagram of the proposed system.

II. SYSTEM DESCRIPTION

The proposed pedestrian detection and tracking system (Fig. 1) consists of four stages: motion detection, pedestrian detection, pedestrian tracking, and re-identification.

A. Motion Detection

In the first stage, motion detection is performed by means of a background segmentation algorithm founded upon pixel-wise Gaussian mixture models. The aim of this step is to effectively determine potential regions containing moving pedestrians to guide and, therefore, reduce the computational cost of the subsequent detection process. Potential misdetections of static pedestrians at this stage are recovered by a full-image (i.e., unguided) detection performed every several frames.

B. Part-body Detection

To solve the problem raised by partial occlusions, we introduce part-body detectors as a supplement to a whole-body detector based on HOG features. The purpose of the part-body detectors is twofold: if a whole-body detection is present, they enhance the confidence in pedestrian detection; if it is not present, a pedestrian detection is still generated from several part-body detections unless it consists of only one or two parts with low confidence, in which case they are filtered, thus

becoming misdetections. In our system, an upper body detector based on Haar and HOG features is trained via Adaboost [6].

C. Compressive Tracking

Once detections are obtained, compressive tracking is applied to estimate the position of pedestrians in the next frame. Specifically, we use local statistical features to construct the appearance model, in which non-adaptive random projections are computed to acquire a sparse matrix that represents the feature space of the tracked objects. Using this framework for both tracked boxes and background, we train a binary Bayes classifier to find the region nearby with the highest confidence as the tracking result. Moreover, the classifier is updated on every frame, acting as an online updating algorithm with an outstanding real-time performance and high accuracy, efficiency, and robustness.

D. Re-identification

To improve object (i.e., pedestrian) labeling and identification, we implement re-identification using fingerprints created by a perceptual hash algorithm [7]. Such module is able to recognize an existing pedestrian in a databased of tracked objects, providing long-term consistent pedestrian labeling.

III. EXPERIMENTAL RESULTS

We tested the proposed model in both indoor and outdoor scenes using a PTZ IP camera. For the detection stage, we used the miss rate metric in [8] to describe the performance of the detector. The combination of whole-body detector and part-body detector resulted in a significant reduction of the miss rate with respect to that of the former alone (see Fig. 2.a, “part-body detections” label). When motion detection is used to guide the pedestrian detection stage, the amount of false positives per image decreases further.

The performance of the tracker was evaluated, using as ground truth that provided by the Video Annotation Tool in [9], and is reported in the precision plot [10] of Fig. 2.b. The precision increases with the location error threshold, but most of the tracked bounding boxes are closer than 20 pixels (approx. 6% of the image width) from the ground truth ones. Some representative results obtained with the proposed strategy are illustrated in Fig. 3 for an indoor sequence with four moving objects (left) and an outdoor scenario with two moving objects (right).

The system has proven to be able to run in real-time conditions at five frames per second, without code optimization. The mean times required by each stage in the system are shown in Table I.

IV. CONCLUSION

In this paper, we propose a novel system for pedestrian detection and tracking in real-time situations. The main contribution of this system is the combination of multiple body

detectors and a strategy for compressive tracking. Part-body detectors strengthen the results of the whole-body detector and reduce the miss rate, while the compressive tracking takes advantage of local statistical features to efficiently describe a candidate pedestrian. Promising results have been obtained.

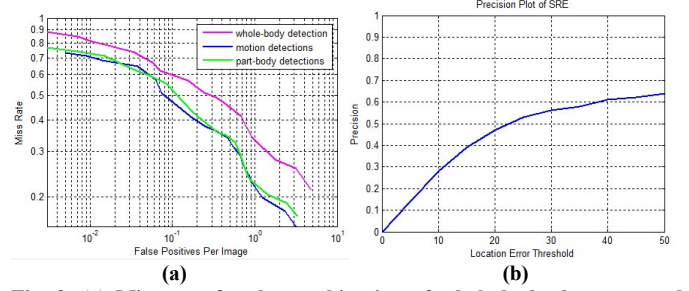


Fig. 2. (a) Miss rate for the combination of whole-body detectors and part-body detectors. Precision plot of the compressive tracker (b).



Fig. 3. Results on real-time test sequences in both indoor (left) and outdoor (right) conditions with IP Network Cameras.

TABLE I
TIME COST OF EACH SYSTEM STAGE IN MS.

Stage	Acquisit.	Motion Detect.	Body Detect.	Tracking	Re-Identif.
Time(ms)	1.155	98.242	38.893	32.949	12.629

REFERENCES

- [1] L. Wang, W. Hu, T. Tan, “Recent developments in human motion analysis,” *Pattern recognition*, 2003, 36(3): 585-601.
- [2] N. Dalal, B. Triggs, “Histograms of oriented gradients for human detection,” *IEEE Conf. Computer Vision and Pattern Recognition*, 2005, 1: 886-893.
- [3] G. R. Bradski, “Real time face and object tracking as a component of a perceptual user interface,” *Fourth IEEE Workshop on Applications of Computer Vision*, 1998: 214-219.
- [4] Z. Zivkovic, “Improved adaptive Gaussian mixture model for background subtraction,” *Proc. IEEE Int. Conf. Pattern Recognition*, 2004, 2: 28-31.
- [5] K. Zhang, L. Zhang, M.H. Yang, “Real-time compressive tracking,” *European Conf. Computer Vision*, Springer, 2012, 864-877.
- [6] Y. Freund, R. E. Schapire, “A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting,” *J. Computer and System Sciences*, 1997, 55(1):119-139.
- [7] B. Yang, F. Gu, X. Niu, “Block mean value based image perceptual hashing,” *Proc. IEEE Int. Conf. Intelligent Information Hiding and Multimedia Signal Processing*, 2006, 167-172.
- [8] P. Dollár, S. Belongie, P. Perona, “The Fastest Pedestrian Detector in the West,” *British Machine Vision Conf.*, 2010, 68.1-68.11.
- [9] C. Vondrick, D. Patterson, D. Ramanan, “Efficiently Scaling Up Crowdsourced Video Annotation,” *Int. J. Computer Vision*, 2013, 101(1): 184-204.
- [10] Y. Wu, J. Lim, M. H. Yang, “Online object tracking: A benchmark,” *IEEE Conf. Computer Vision and Pattern Recognition*, 2013, 2411-2418.